Random Forest の寄与率を用いた効率的な特徴選択法の提案

三品 陽平† 嶋崎 克也† 藤吉 弘亘†

†中部大学

E-mail: mishi@vision.cs.chubu.ac.jp, hf@cs.chubu.ac.jp

Abstract

画像認識の問題では，識別器を構築する際に高次元ベクトルの特徴量が用いられることが多い．高次元な特徴量を用いると，学習速度の低下やモデルの可読性を低下させる原因となる．そのため，識別に有効な特徴次元を選択する手法として，Sequential Backward Selection(SBS) 法が用いられている．しかし，ラッパー法である SBS 法は特徴選択する際に識別器の構築と評価を繰り返すために膨大な時間を要するという問題がある．そこで，本研究では Random Forest における特徴次元の寄与率を定義し，寄与率を用いた効率的な特徴選択法を選択する．評価実験より，提案手法は従来法の SBS 法と比較し同等の削減率において，特徴選択時間を大幅に削減することができた．

1 はじめに

計算機の発展により大規模なデータを高速に処理することが可能となり，コンピュータビジョンやパターン認識の分野においても，識別器を構築するために大規模なデータを用いる傾向がある [1]．一般に画像認識の問題に対しては，高次元な特徴量を用いて識別器を構築することが多い．高次元な特徴量の中には識別性能に寄与しない特徴次元が存在する場合があり，学習速度や学習モデルの可読性を低下させる要因となる．また，このような高次元特徴量を用いて識別時に大量のメモリを消費し，演算コストが増加するという問題がある．そのため，特徴量を選択する手法が提案されている．

従来の特徴選択法は，エントロピーなどの評価関数を基に選択をするフィルタや汎化誤差などにより選択するラッパー法などに分けられる．フィルタは，学習サンプルを用いて評価関数により選択するため，高速に処理できる．しかし，目的のデータセットごとに最適な評価関数の設計が必要となる．ラッパー法は，特徴量の部分集合から識別器を構築し，識別性能を評価する．部分集合の組み合わせを変化させ識別器を構築する．識別器の汎化性能を評価し，最も性能の高い識別器を構築した部分集合を識別性能の高い特徴量として選択する．ラッパー法は，汎化性能により特徴量を選択しているため，フィルタよりも識別性能を維持したまま特徴次元を削減できる．しかし，部分集合毎に識別器を構築する必要があるため選択時間が膨大になる．

そこで，本研究では識別性能を維持してラッパー法のように識別器の構築と評価を繰り返す必要のない効率的な特徴選択を目的とする．本研究では識別器に，コンピュータビジョンやパターン認識の分野で注目されている Random Forest[2] を用いる．Random Forest は，複数の決定木を構築する決定木学習，アンサンブル学習を組み合わせた学習アルゴリズムである．また，ランダム学習を取り入れることで，ノイズに頑健で高速な学習が特徴である．本研究では，構築された決定木において各特徴次元が識別性能に寄与している度合いを寄与率として定義し，寄与率を基に特徴次元を削減する．寄与率は，識別器を一度構築すれば算出できるため，繰り返し識別器を構築するラッパー法よりも高速に特徴選択が可能となる．

2 従来の特徴選択法

識別能力の高い特徴次元の組み合わせを選択することで，識別器の可読性の改善，識別時の特徴抽出時間の短縮, メモリ消費量の削減に貢献する．特徴選択において解くべき最適化問題は，K 個の特徴次元の中から汎化性能が高く，最も削減率の高い d 個の特徴次元を選択することである．特徴選択法は，大きく分けてフィルタとラッパー法の二種類に分類される．フィルタは，学習サンプルを用いてエントロピーなどの評価関数により，特徴次元を選択するため高速に処理できる．しかし，データセットや学習モデルに合わせて最適な評価関数を選択する必要がある．ラッパー法は，特徴量の部分集合から識別器を構築し，識別性能を評価する．部分集合の組み合わせを変化させ識別器を構築する．識別器の汎化性能を評価し，最も性能の高い識別器を構築した部分集合を識別性能の高い特徴量として選択する．

2.1 ラッパー法

ラッパー法は，汎化性能により評価しているためフィルタに比べて削減率は高い．この最適化問題を解くラッパー法の代表的な手法として，総当たり法，SFS 法，SBS 法，プラス s マイナス r 法がある．以下にこれらの手法の概要を述べる．

2.1.1 総当たり法

総当たり法は全探索法ともよばれ，${}_KC_d$ 個の全ての組み合わせを探索する方法である．特徴選択法の中で，最適解を得ることが唯一保証されているが，特徴次元の数 K と選択する次元数 d の値が大きくなるにつれて計算量は爆発的に増加する問題がある．このため，特徴次元の数が大きい場合は実用的ではない．

2.1.2 Sequential Forward Selection 法

Whiteney が提案した逐次型前向き最適化法 (Sequential Forward Selection:SFS)[4] がある．この SFS 法は，各特徴次元で識別器を構築し，最も識別率が高い識別器の特徴次元を第 1 特徴次元として選択する．そして，すでに選択した特徴次元と組み合わせて識別器を構築し，識別率が最も高い識別器の特徴次元を 1 つ追加し第 2 特徴次元とする．第 3 特徴次元以降も同様に探索を行い，d 個になるまで逐次追加していく方法である．ラッパー法の中では計算コストが低いため，実行可能な方法として用いられている．

2.1.3 Sequential Backward Selection 法

SFS 法と対照的な，逐次型後向き最適化法 (Sequential Backward Selection:SBS)[3] が Marill により提案されている．SBS 法は，特徴次元を 1 次元除いた部分集合ごとに識別器を構築し，誤識別率などの評価値を評価する．評価値が最も高い識別器を構築した際に，除いていた特徴次元は識別性能に寄与しない次元と考え削除していく方法である．図 1 に SBS 法の流れ図を示す．

2.1.4 プラス s マイナス r 法

Stearns が提案したプラス s マイナス r 法 [5] は，SFS 法と SBS 法を交互に用いる最適化手法で，一度選択した特徴次元を削除することができる．k 個の特徴次元が既に選択され，選択された特徴次元の集合 X_k が与えられたとする．まず，X_{k+s} を得るため，X_k に SFS 法を s 回適用し，特徴次元を選択する．その後，X_{k+s-r} となるまで SBS 法により特徴次元を削減する．目的の次元数である d となるまで SFS 法と SBS 法を繰り返し適用する．通常は $s > r$ で，例えば $s = 2, r = 1$ としてプラス s マイナス r 法を用いる．

2.2 ラッパー法の問題点

ラッパー法は，識別性能を評価し特徴次元を選択しているため，フィルタと比べデータセットや学習モデルに最適化する必要がなく安定して特徴選択できる．しかし，特徴次元を選択する際に次元数分の識別器を学習する必要がある．特徴選択するために，識別器の学習と評価を繰り返す必要があるため，特徴次元の増加に伴い処理時間も膨大に増加する．

3 Random Forest

本章では，寄与率を用いた特徴選択法に利用する Random Forest について述べる．Random Forest は，複数の決定木構造を持つマルチクラス識別器を構築する学習アルゴリズムである．Random Forest のアルゴリズムの特徴は，Bagging[7] と同様にブートストラップを取り入れ過学習を防ぐ点，Random Feature Selection[8] を取り入れ特徴ベクトルの次元数が大きくても高速に学習が可能な点である．このようなメリットを持つため，コンピュータビジョンの分野でも，セマンティックセグメンテーション [9]，文字認識 [10]，物体認識 [11][12]，人体姿勢推定 [13] で用いられている．

3.1 学習

Random Forest は，学習サンプルからサブセットを作成し，複数の決定木構造を持つ識別器を構築する．各決定木は，分岐ノードと末端ノードにより構成され，分岐ノードを繰り返し作成し，一定の基準により分岐が不可能になった際に，末端ノードを作成する．学習アルゴリズムを Algorithm1 に示す．木の数 T，木の最大の深さ D とする．学習サンプル集合 $\mathcal{I} = \{\mathbf{x}_1, y_1\}, \ldots, \{\mathbf{x}_N, y_N\}, \mathbf{x}_i \in \mathcal{X}, y_i \in \{1, 2, \ldots, C\}$ を入力し，サブセットを作成する．サブセットは学習サンプル $\mathcal{I}$ からサンプルの重複を許容してランダムに選択する．サブセット一つを用いて決定木を構築する．決定木の分岐ノードはある特徴量 f_k としきい値 τ_h を用いて左もしくは右へサンプルを分岐させる分岐関数が保存されている．分岐関数は，特徴選択回数 K としきい値選択回数 H から，候補をランダムに $K \times H$ 個選択し，情報利得 ΔE が最も高い候補に決定する．ある分岐ノード n にたどり着いたサンプル集合 $\mathcal{I}_n$，分岐関数の候補 f_k と τ_h により $\mathcal{I}_l$ と $\mathcal{I}_r$ に分割する．この I_l と I_r を用いて，式 (1) により情報利得 ΔE を算出する．情報利得は，現在のノードのエントロピーから子ノードのエントロピーの和を引いたものであり，分岐関数によりどの程度情報量が減少したかを表している．子ノードのエントロピーが小さくなると情報利得は大き

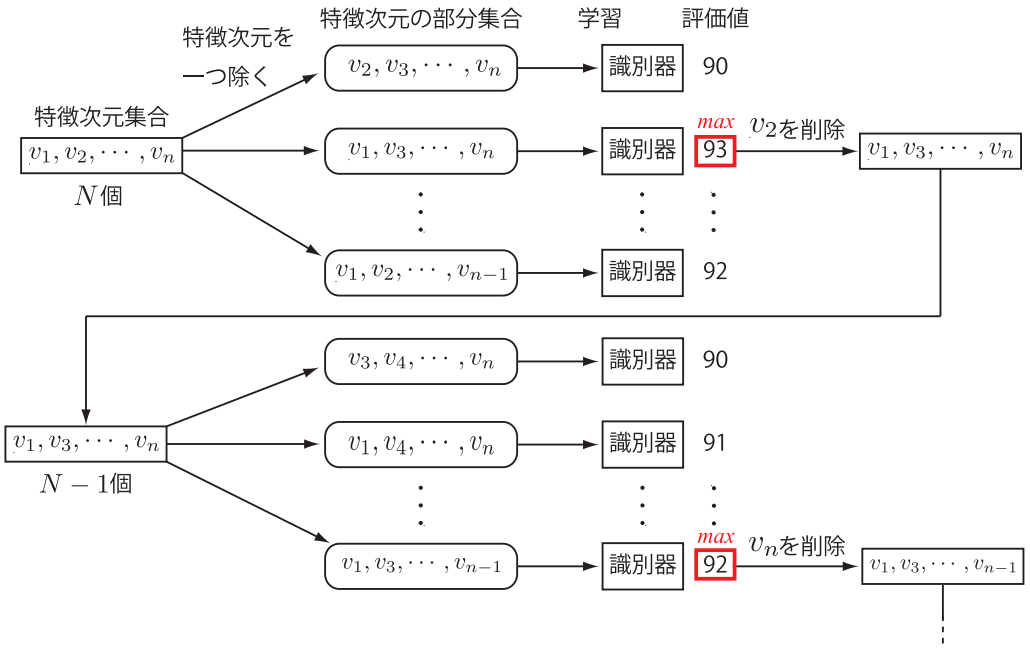

図 1　Sequential Backward Selection(SBS) 法の流れ

Algorithm 1 Random Forest の学習アルゴリズム

Require: 学習サンプル $\{\mathbf{x}_1, y_1\}, \ldots, \{\mathbf{x}_N, y_N\}$;
$\mathbf{x}_i \in \mathcal{X}, y_i \in \{1, 2, \ldots, M\}$

Run:
  for $t = 1 : T$ **do**
    学習サンプルからサブセット $\mathcal{I}_t$ を作成.
    $\Delta E_{max} \Leftarrow -\infty$.
    for $k = 1 : K$ **do**
      特徴量からランダムに f_k を 1 次元選択する.
      for $h = 1 : H$ **do**
        しきい値 τ_h をランダムに選択する.
        f_k と τ_h によりサンプル集合 $\mathcal{I}_n$ を $\mathcal{I}_l$ と $\mathcal{I}_r$ に分割.
        情報利得 ΔE を算出:
        $\Delta E = E(\mathcal{I}_n) - \frac{|\mathcal{I}_l|}{|\mathcal{I}_n|} E(\mathcal{I}_l) - \frac{|\mathcal{I}_r|}{|\mathcal{I}_n|} E(\mathcal{I}_r)$.
        if $\Delta E > \Delta E_{max}$ **then**
          $\Delta E_{max} \Leftarrow \Delta E$
        end if
      end for
    end for
    if $\Delta E_{max} = 0$ または最大の深さ D に達した **then**
      末端ノードにクラス確率 $P(c|l)$ を保存.
    else
      分岐を続ける.
    end if
  end for

くなり，カテゴリをよく分割する分岐関数と表される．

$$\Delta E = E(\mathcal{I}_n) - \frac{|\mathcal{I}_l|}{|\mathcal{I}_n|} E(\mathcal{I}_l) - \frac{|\mathcal{I}_r|}{|\mathcal{I}_n|} E(\mathcal{I}_r) \qquad (1)$$

ここで，関数 $E(I)$ は情報エントロピーを表し式 (2) により算出される．

$$E(\mathcal{I}) = -\sum_{i=1}^{n} p(c_i) \log p(c_i) \qquad (2)$$

ここで，$p(c_i)$ はクラス c_i の確率を表しており，学習サンプルの教師信号の出現頻度により求められる．これらの処理を繰り返すことによりサンプル集合を分割し，情報利得が 0 になった場合や最大の深さ D に達した場合に末端ノード l を作成し，到達したサンプル集合から各カテゴリの出現確率 $P(c|l)$ を計算する．このように，各決定木が構築される．

3.2 識別

Random Forest の識別のアルゴリズムについて説明する．未知入力サンプル $\mathbf{x}$ をすべての決定木に入力し，分岐関数により左右に分岐させ決定木をトラバーサルする．そして，たどり着いた末端ノードに保存されているクラス確率 $P_t(c|\mathbf{x})$ を出力する．式 (3) に示すように，すべての決定木の出力の平均を算出する．

$$P(c|\mathbf{x}) = \frac{1}{T} \sum_{t=1}^{T} P_t(c|\mathbf{x}). \qquad (3)$$

最終出力 $\hat{y}$ は，式 (4) により最も確率の高いクラスに判別する．

$$\hat{y} = \arg\max_{c} P(c|\mathbf{x}). \qquad (4)$$

4 提案手法

本研究は，識別器の構築と評価を繰り返す必要のない効率的な特徴選択を提案する．提案手法は，構築された識別器の識別性能に各特徴次元が寄与している割合を評価基準とし，特徴選択を行う．寄与率は，Random Forest を一度構築すれば算出できるため，繰り返し識別器を構築するラッパー法よりも高速に特徴選択が可

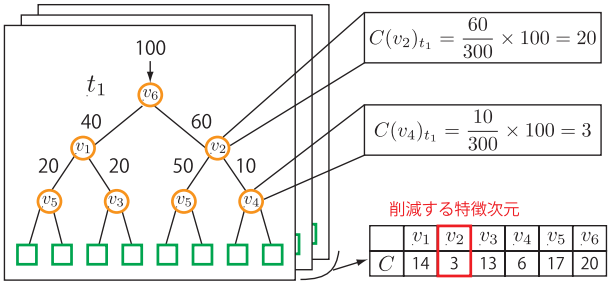

図 2　寄与率の算出方法

能となる．寄与率を用いた特徴選択法は，逐次算出型と一括算出型を提案する．逐次算出型は，Backward 方式に 1 次元づつ特徴次元を削減する場合に，1 次元削減後に寄与率を算出する．一括算出型は，全特徴量を用いて Random Forest を構築した場合の寄与率を算出し，その寄与率により特徴次元を選択する．

4.1　寄与率の算出

図 2 に寄与率の算出方法を示す．決定木の上層の分岐ノードで選ばれた特徴次元は，多くの学習サンプルを分割しているため，識別器の識別能力への寄与が高いと考えられる．そのため，Random Forests における寄与率を分岐ノードに辿り着いたサンプル数を用いて，寄与率 $C(i)$ を式 (5) と定義する．

$$C(i) = \frac{1}{T}\sum_{t=1}^{T}\frac{\sum_j^J S_{t,j}\delta[f_j,i]}{\sum_j^J S_{t,j}} \times 100 \qquad (5)$$

ここで，$\delta[f_j,i]$ はデルタ関数を表し，ノード j において選択された次元番号が i と同じ場合は 1 を返し，それ以外は 0 を返す．よって，分子は i 番目の特徴次元が選ばれた分岐ノード j が分割したサンプル数 $S_{t,j}$ の総数である．分母は，各決定木の全ての分岐ノードが分割したサンプル数 $S_{t,j}$ を示す．構築した木 T 本から寄与率を算出し，平均をとることで各次元の寄与率を算出する．

4.2　寄与率を用いた逐次算出による特徴選択法

寄与率を用いた逐次算出型の特徴選択法を Algorithm2 に示す．提案手法は，まず Random Forests を構築し寄与率を算出する．構築した識別器の評価を行い終了判定をする．条件を満たし削減を続ける場合は，寄与率の最も低い特徴次元を削除する．図 3 に逐次算出による特徴選択法の流れを示す．

4.3　寄与率を用いた一括算出による特徴選択法

寄与率の一括算出による特徴選択アルゴリズムを Algorithm3 に示す．図 4 に一括算出による特徴選択の流れ図をしめす．一括算出型は，すべての特徴次元から算出した寄与率に基づき削減する手法である．一括算出型は，Random Forests を構築する毎に寄与率を算出しないため，逐次算出型より計算コストが少ない．

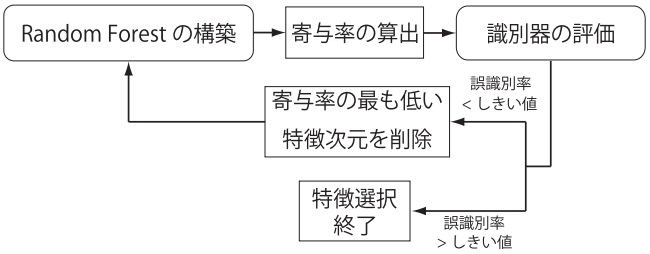

図 3　逐次算出による特徴選択法の流れ

Algorithm 2 寄与率の逐次算出による特徴選択法

Require: 木の数：T

Run:

for $k = 1 : K$ **do**
　現在の特徴量を用いて RF を構築
　for $i = 0 : I^{(k)}$ **do**
　　構築後の RF から各特徴次元の寄与率 $C(i)$ を算出:
　　$C(i) = \frac{1}{T}\sum_{t=1}^{T}\frac{\sum_j^J S_{t,j}\delta[f_j,i]}{\sum_j^J S_{t,j}} \times 100$
　end for
　if 終了判定の条件を満す場合 **then**
　　寄与率の最も低い特徴次元を削減.
　else
　　特徴選択を終了.
　end if
end for

5　評価実験

提案手法の有効性を示すため，SBS 法と比較実験を行う．実験の終了条件は，全ての特徴次元を用いて学習，評価をした際の誤識別率から 10%増加するまでとする．評価方法は，3-hold cross-validation による汎化誤差を測定する．

5.1　Dataset

評価実験には UCI Machine Learning Repository[14] のデータセットを用いる．UCI Machine Learning

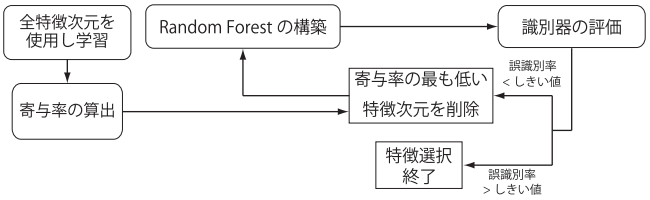

図 4　一括算出による特徴選択法の流れ

Algorithm 3 寄与率の一括算出による特徴選択法

Require: 木の数：T

Run:

すべての特徴量を用いて RF を構築

for $i = 0 : I^{(k)}$ **do**

構築後の RF から各特徴次元の寄与率 $C(i)$ を算出:

$$C(i) = \frac{1}{T}\sum_{t=1}^{T}\frac{\sum_j^J S_{t,j}\delta[f_j,i]}{\sum_j^J S_{t,j}} \times 100$$

end for

for $k = 1 : K$ **do**

if 終了判定の条件を満す場合 **then**

寄与率の最も低い特徴次元を削減.

else

特徴選択を終了.

end if

end for

表 1　データセットの概要

データセット名	サンプル数	クラス数	特徴次元数
Pendigits	14988	10	16
Waveform	5000	3	21
Spambase	4601	2	57
Optdigits	5620	10	64

Repository はカリフォルニア大学アーバイン校で公開されている機械学習アルゴリズムのベンチマークデータセットである．今回の実験では，UCI Machine Learning Repository の中の Pendigits, Waveform, Spambase, Optdigits の 4 つのデータセットを使用する．データセットの概要を表 1 に示す．

5.2 学習パラメータ

実験に用いた Random Forests の学習パラメータを表 2 に示す．学習パラメータは，全ての特徴次元を用いた深さを 10〜30，木の数を 10〜200 まで 10 刻みに変化させ学習，評価の結果をから，誤差が最小となるパラメータを用いた．また，各データセットで固定なパラメータは，特徴選択回数を各特徴次元数の二乗根とし，しきい値選択回数を 10 回，サブセットの割合を 1.0 に設定した．

5.3 実験結果

図 5，6，7，8 に各データセットの削減率に対する誤識別率を示す．図より提案手法の誤識別率は，従来法と削減率が等しいとき同等の誤識別率で特徴選択ができる．逐次算出型は，一括算出型よりも特徴量の変動を考慮し評価するため，削減率が高いと考えられる．

図 9 に従来法と選択した特徴次元の共通率を示す．ま

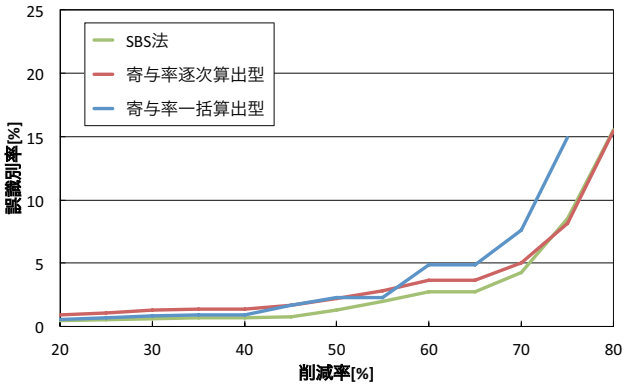

図 5　Pendigits の削減率に対する誤識別率

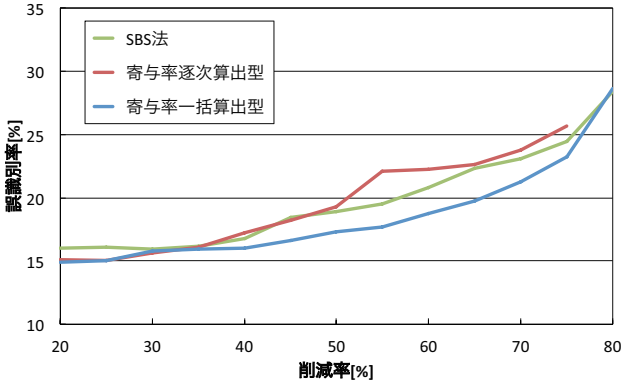

図 6　Waveform の削減率に対する誤識別率

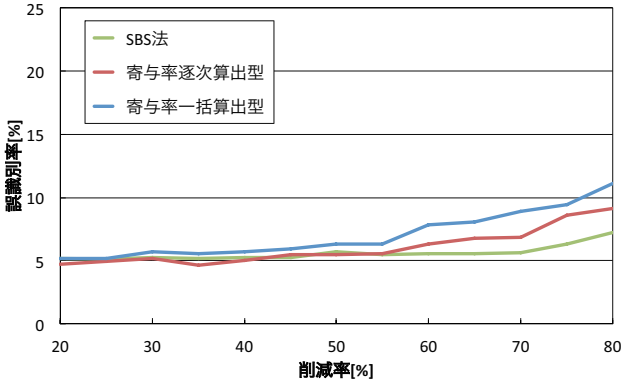

図 7　Spambase の削減率に対する誤識別率

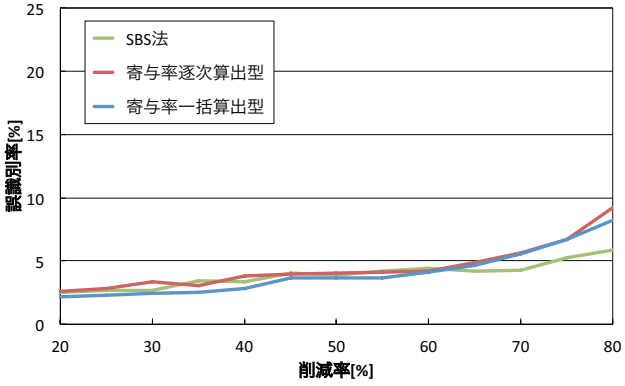

図 8　Optdigits の削減率に対する誤識別率

表 2　Random Forests の学習パラメータ

データセット名	深さ	木の数
Pendigits	20	80
Waveform	30	150
Spambase	30	90
Optdigits	20	110

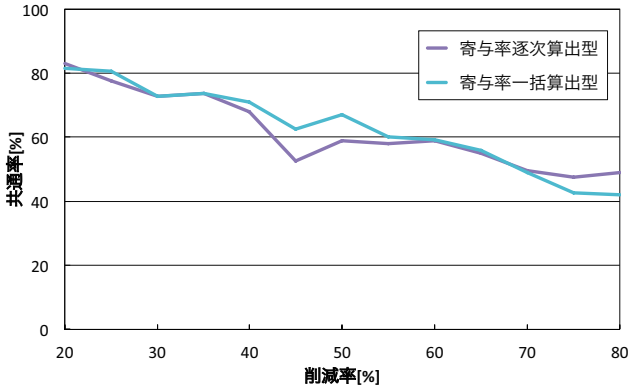

図 9　SBS 法と提案手法の選択した特徴の共通率

た，SBS 法で選択された特徴次元と提案手法により選択された特徴次元の共通率は，削減するごとに低下するが特徴次元を 80%削減した時点でも約 5 割が従来法と同じ特徴次元を選択している．図 10 にデータセット毎の特徴次元 1 つを削減するための処理時間を示す．提案手法は従来手法と比較して，特徴選択時間を大幅に短縮できた．従来法は，特徴次元数分の識別器を学習する必要があるため，特徴次元数を N とすると $\mathcal{O}(N)$ 回の識別器を構築する必要がある．提案手法では，一つの識別器から寄与率を算出できるため，識別器の構築は $\mathcal{O}(1)$ 回であるため選択時間が短縮できた．そのため，特徴次元数の高いデータセットほど選択時間の差が大きい結果となった．

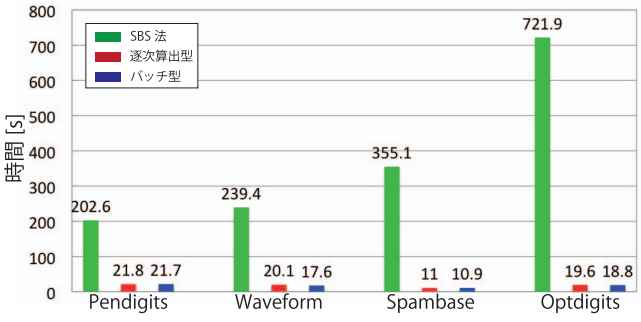

図 10　特徴選択に要する時間

6　おわりに

本稿では，Random Forest の決定木から各特徴次元が識別性能に寄与する割合を寄与率と定義し，寄与率をもとに特徴選択する方法を提案した．評価実験の結果より，提案手法は従来法である SBS 法と比較し，同等の削減率を示した．また，従来の SBS 法は特徴次元数 N のとき $\mathcal{O}(N)$ 回だけ識別器を構築する必要がある．それに対して，提案手法は $\mathcal{O}(1)$ 回しか識別器を構築しないため，特徴選択にかかる時間を大幅に短縮した．今後は，決定木の構造を利用して特徴次元間の共起性も評価することで，最適な特徴次元の組み合わせを選択する方法を検討する．

参考文献

[1] 柳井啓司，“一般物体認識の現状と今後”，情報処理学会論文誌，コンピュータビジョンとイメージメディア，vol. 48, no. 16, pp. 1-24, 2007.
[2] L. Breiman,“Random Forests”, Machine Learning, vol. 45, pp.5-32, 2001.
[3] Marill, T, D. M. Green,“On the effectiveness of receptors in recognition system,”IEEE Trans. Inform. Theory 9, pp. 11-17, 1963.
[4] Whitney, A. W,“A direct method of nonparametric measurement selection,” IEEE Trans. Comput.20 ,pp. 1100-1103, 1997.
[5] S. D. Stearns: On selecting features for pattern classifies, Proc. Third Internat. Conf. Pattern Recognition, pp. 71-75, 1976.
[6] P. Pudil, J. Novovicora and J. Kittler: Floating search methods in feature selection, Pattern Recognition Letters, Vol. 15, No. 11, pp. 1119-1125, 1994.
[7] L. Breiman, “Bagging Predictors”, Machine Learning, vol. 24, no. 2, pp. 123-140, 1996.
[8] Ho, T. K.,“The random subspace method for constructing decision forests”, IEEE Transactions on Pattern Analysisand Machine Intelligence, vol.20 no.8, pp.832-844, 1998.
[9] J. Shotton, M. Johnson and R. Cipolla, “Semantic texton forests for image categorization and segmentation”, Computer Vision and Pattern Recognition, 2008.
[10] Y. Amit and D. Geman,“Shape Quantization and Recognition with Randomized Trees”, Neural Computation, No.9, pp.1545-1588, 1997.
[11] V. Lepetit and p. Fua,“Keypoint recognition using randomized trees”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 9, pp.1465-1479, 2006.
[12] Gall, J. and Yao, A. and Razavi, N. and Van Gool, L. and Lempitsky, V.,“Hough forests for object detection, tracking, and action recognition”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.33, no.11, pp.2188-2202, 2011.
[13] J. Shotton,and A. Fitzgibbon, and Cook, M. and Sharp, T. and Finocchio, M. and Moore, R. and Kipman, A. and Blake, A.,“Real-time human pose recognition in parts from single depth images”, Computer Vision and Pattern Recognition, 2011.
[14] UCI Machine Learning Repository, http://archive.ics.uci.edu/ml/.